CelebA-Spoof: Large-Scale Face Anti-spoofing Dataset with Rich Annotations

Yuanhan Zhang[1], ZhenFei Yin[2], Yidong Li[1], Guojun Yin[2], Junjie Yan[2], Jing Shao[2(✉)], and Ziwei Liu[3]

[1] Beijing Jiaotong University, Beijing, China
{18120454,yidongli}@bjtu.edu.cn
[2] SenseTime Group Limited, Hong Kong, China
{yinzhenfei,yinguojun,yanjunjie,shaojing}@sensetime.com
[3] The Chinese University of Hong Kong, Hong Kong, China
zwliu@ie.cuhk.edu.hk

Abstract. As facial interaction systems are prevalently deployed, security and reliability of these systems become a critical issue, with substantial research efforts devoted. Among them, face anti-spoofing emerges as an important area, whose objective is to identify whether a presented face is live or spoof. Though promising progress has been achieved, existing works still have difficulty in handling complex spoof attacks and generalizing to real-world scenarios. The main reason is that current face anti-spoofing datasets are limited in both quantity and diversity. To overcome these obstacles, we contribute a large-scale face anti-spoofing dataset, **CelebA-Spoof**, with the following appealing properties: *1) Quantity:* CelebA-Spoof comprises of 625,537 pictures of 10,177 subjects, significantly larger than the existing datasets. *2) Diversity:* The spoof images are captured from 8 scenes (2 environments * 4 illumination conditions) with more than 10 sensors. *3) Annotation Richness:* CelebA-Spoof contains 10 spoof type annotations, as well as the 40 attribute annotations inherited from the original CelebA dataset. Equipped with CelebA-Spoof, we carefully benchmark existing methods in a unified multi-task framework, **Auxiliary Information Embedding Network (AENet)**, and reveal several valuable observations. Our key insight is that, compared with the commonly-used binary supervision or mid-level geometric representations, rich semantic annotations as auxiliary tasks can greatly boost the performance and generalizability of face anti-spoofing across a wide range of spoof attacks. Through comprehensive studies, we show that CelebA-Spoof serves as an effective training data source. Models trained on CelebA-Spoof (without fine-tuning) exhibit state-of-the-art performance on standard benchmarks such as CASIA-MFSD. The datasets are available at https://github.com/Davidzhangyuanhan/CelebA-Spoof.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-58610-2_5) contains supplementary material, which is available to authorized users.

A. Vedaldi et al. (Eds.): ECCV 2020, LNCS 12357, pp. 70–85, 2020.
https://doi.org/10.1007/978-3-030-58610-2_5

Keywords: Face anti-spoofing · Large-scale dataset

1 Introduction

Face anti-spoofing is an important task in computer vision, which aims to facilitate facial interaction systems to determine whether a presented face is live or spoof. With the successful deployments in phone unlock, access control and e-wallet payment, facial interaction systems already become an integral part in the real world. However, there exists a vital threat to these face interaction systems. Imagine a scenario where an attacker with a photo or video of you can unlock your phone and even pay his bill using your e-wallet. To this end, face anti-spoofing has emerged as a crucial technique to protect our privacy and property from being illegally used by others.

Most modern face anti-spoofing methods [8,14,31] are fueled by the availability of face anti-spoofing datasets [4,5,18,24,29,32,34], as shown in Table 1. However, there are several limitations with the existing datasets: 1) *Lack of Diversity.* Existing datasets suffer from lacking sufficient subjects, sessions and input sensors (*e.g.* mostly less than 2000 subject, 4 sessions and 10 input sensors). 2) *Lack of Annotations.* Existing datasets have only annotated the type of spoof type. Face anti-spoof community lacks a densely annotated dataset covering rich attributes, which can further help researchers to explore face anti-spoofing task with diverse attributes. 3) *Performance Saturation.* The classification performance on several face anti-spoofing datasets has already saturated, failing to evaluate the capability of existing and future algorithms. For example, the recall under FPR = 0.5% on SiW and Oulu-NPU datasets using vanilla ResNet-18 has already reached 100.0% and 99.0%, respectively (Fig. 1).

To address these shortcomings in existing face anti-spoofing dataset, in this work we propose a large-scale and densely annotated dataset, **CelebA-Spoof**. Besides the standard *Spoof Type* annotation, CelebA-Spoof also contains annotations for *Illumination Condition* and *Environment*, which express more information in face anti-spoofing, compared to categorical label like *Live/Spoof*. Essentially, these dense annotations describe images by answering questions like "Is the people in the image Live or Spoof?", "What kind of spoof type is this?", "What kind of illumination condition is this?" and "What kind of environment in the background?". Specifically, all live images in CelebA-Spoof are selected from CelebA [20], and all Spoof images are collected and annotated by skillful annotators. CelebA-Spoof has several appealing properties. **1) Large-Scale.** CelebA-Spoof comprises of a total of 10177 subjects, 625537 images, which is the largest dataset in face anti-spoofing. **2) Diversity.** For collecting images, we use more than 10 different input tensors, including phones, pads and personal computers (PC). Besides, we cover images in 8 different sessions. **3) Rich Annotations.** Each image in CelebA-Spoof is defined with 43 different attributes: 40 types of *Face Attribute* defined in CelebA [20] plus 3 attributes of face anti-spoofing, including: *Spoof Type*, *Illumination Condition* and *Environment*. With

Fig. 1. A quick glance of CelebA-Spoof face anti-spoofing dataset with its attributes. Hypothetical space of scenes are partitioned by attributes and Live/Spoof. In reality, this space is much higher dimensional and there are no clean boundaries between attributes presence and absence

rich annotations, we can comprehensively investigate face anti-spoofing task from various perspectives.

Equipped with CelebA-Spoof, we design a simple yet powerful network named **A**uxiliary information **E**mbedding **N**etwork (**AENet**), and carefully benchmark existing methods within this unified multi-task framework. Several valuable observations are revealed: **1)** We analyze the effectiveness of auxiliary **geometric information** for different spoof types and illustrate the sensitivity of geometric information to special illumination conditions. Geometric information includes *depth map* and *reflection map*. **2)** We validate auxiliary **semantic information**, including face attribute and spoof type, plays an important role in improving classification performance. **3)** We build three CelebA-Spoof benchmarks based on this two auxiliary information. Through extensive experiments, we demonstrate that our large-scale and densely annotated dataset serves as an effective data source in face anti-spoofing to achieve state-of-the-art performance. Furthermore, models trained with auxiliary semantic information exhibit great generalizability compared to other alternatives.

In summary, the **contributions** of this work are three-fold:**1)** We contribute a large-scale face anti-spoofing dataset, **CelebA-Spoof**, with 625,537 images from 10,177 subjects, which includes 43 rich attributes on face, illumination, environment and spoof types. **2)** Based on these rich attributes, we further propose a simple yet powerful multi-task framework, namely **AENet**. Through AENet, we conduct extensive experiments to explore the roles of semantic information and geometric information in face anti-spoofing. **3)** To support comprehensive evaluation and diagnosis, we establish three versatile benchmarks to evaluate the performance and generalization ability of various methods under different carefully-designed protocols. With several valuable observations revealed,

Table 1. The comparison of CelebA-Spoof with existing datasets of face anti-spoofing. Different illumination conditions and environments make up different sessions, (V means video, I means image; Ill. Illumination condition, Env. Environment; - means this information is not annotated)

Dataset	Year	Modality	#Subjects	#Data(V/I)	#Sensor	#Semantic Attribute		
						#Face Attribute	Spoof type	#Session (Ill.,Env.)
Replay-Attack [5]	2012	RGB	50	1,200 (V)	2		1 Print, 2 Replay	1 (-,-)
CASIA-MFSD [35]	2012	RGB	50	600 (V)	3		1 Print, 1 Replay	3 (-,-)
3DMAD [7]	2014	RGB/Depth	14	255 (V)	2		1 3D mask	3 (-,-)
MSU-MFSD [30]	2015	RGB	35	440 (V)	2		1 Print, 2 Replay	1 (-,-)
Msspoof [10]	2015	RGB/IR	21	4,704 (I)	2	\	1 Print	7 (-,7)
HKBU-MARs V2 [17]	2016	RGB	12	1,008 (V)	7		2 3D masks	6 (6,-)
MSU-USSA [25]	2016	RGB	1,140	10,260 (I)	2		2 Print, 6 Replay	1 (-,-)
Oulu-NPU [4]	2017	RGB	55	5,940 (V)	6		2 Print, 2 Replay	3 (-,-)
SiW [19]	2018	RGB	165	4,620 (V)	2		2 Print, 4 Replay	4 (-,-)
CASIA-SURF [33]	2018	RGB/IR/Depth	1,000	21,000 (V)	1		5 Paper Cut	1 (-,-)
CSMAD [1]	2018	RGB/IR/Depth/LWIR	14	246 (V),17 (I)	1		1 silicone mask	4 (4,-)
HKBU-MARs V1+ [16]	2018	RGB	12	180(v)	1		1 3D mask	1 (1,-)
SiW-M [20]	2019	RGB	493	1,628 (V)	4		1 Print, 1 Replay 5 3D Mask, 3 Make Up, 3 Partial	3 (-,-)
CelebA-Spoof	**2020**	**RGB**	**10,177**	**625,537 (I)**	**>10**	**40**	**3 Print, 3 Replay 1 3D, 3 Paper Cut**	**8 (4,2)**

we demonst rate the effectiveness of CelebA-Spoof and its rich attributes which can significantly facilitate future research.

2 Related Work

Face Anti-spoofing Datasets. Face anti-spoofing community mainly has three types of datasets. First, the multi-modal dataset: 3DMAD [7], Msspoof [6], CASIA-SURF [32] and CSMAD [1]. However, since widespread used mobile phones are not equipped with suitable modules, such datasets cannot be widely used in the real scene. Second is the single-modal dataset, such as Replay Attack [5], CASIA-MFSD [34], MSU-MFSD [29], MSU-USSA [24] and HKBU-MARS V2 [16]. But these datasets have been collected for more than three years. With the rapid development of electronic equipment, the acquisition equipment of these datasets is completely outdated and cannot meet the actual needs. SiW [18], Oulu-NPU [4] and HKBU-MAR V1+ [15] are relatively up-to-date. However, the limited number of subjects, spoof types, and environment (Only indoors) in these datasets does not guarantee for the generalization capability required in the real application. Third, SiW-M [19] is mainly used for Zero-Shot face anti-spoofing tasks. CelebA-Spoof datasets have 625537 pictures from 10177 subjects, 8 scenes (2 environments * 4 illumination conditions) with rich annotations. The characteristic of Large-scale and diversity can further fill the gap between face anti-spoofing dataset and real scenes. With rich annotations we can better analyze face anti-spoofing task. All datasets mentioned above are listed in Table 1.

Face Anti-spoofing Methods. In recent years, face anti-spoofing algorithms have seen great progress. Most traditional algorithms focus on handcrafted features, such as LBP [5,21,22,30], HoG [21,25,30] and SURF [2]. Other works also focused on temporal features such as eye-blinking [23,27] and lips motion

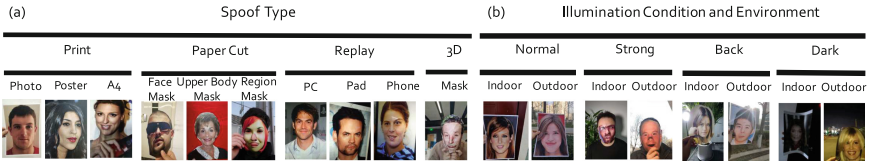

Fig. 2. Representative examples of the semantic attributes (*i.e.* spoof type, illumination and environment) defined upon spoof images. In detail, (a) 4 macro-types and 11 micro-types of spoof type and (b) 4 illumination and 2 types of environmental conditions are defined

[12]. In order to improve the robustness to light changes, some researchers have paid attention to different color spaces, such as HSV [3], YCbcR [2] and Fourier spectrum [13]. With the development of the deep learning model, researchers have also begun to focus on Convolutional Neural Network based methods. [8,14] considered the face PAD problem as binary classification and perform good performance. The method of auxiliary supervision is also used to improve the performance of binary classification supervision. Atoum *et al.* let the full convolutional network to learn the depth map and then assist the binary classification task. Liu *et al.* [15,17] proposed remote toplethysmography (rPPG signal)-based methods to foster the development of 3D face anti-spoofing. Liu *et al.* [18] proposed to leverage depth map combined with rPPG signal as the auxiliary supervision information. Kim *et al.* [11] proposed using depth map and reflection map as the Bipartite auxiliary supervision. Besides, Yang *et al.* [31] proposed to combine the spatial information with the temporal information in the video stream to improve the generalization of the model. Amin *et al.* [10] solved the problem of face anti-spoofing by decomposing a spoof photo into a Live photo and a Spoof noise pattern. These methods mentioned above are prone to over-fitting on the training data, the generalization performance is poor in real scenarios. In order to solve the poor generalization problem, Shao *et al.* [26] adopted transfer learning to further improve performance. Therefore, a more complex face anti-spoofing dataset with large-scale and diversity is necessary. From extensive experiments, CelebA-Spoof has been shown to significantly improve generalization of basic models, In addition, based on auxiliary semantic information method can further achieve better generalization.

3 CelebA-Spoof Dataset

Existing face anti-spoofing datasets cannot satisfy the requirements for real scenario applications. As shown in Table 1, most of them contain fewer than 200 subjects and 5 sessions, meanwhile they are only captured indoor with fewer than 10 types of input sensors. On the contrary, our proposed CelebA-Spoof dataset provides 625, 537 pictures and 10, 177 subjects, therefore offering a superior comprehensive dataset for the area of face anti-spoofing. Furthermore, each image

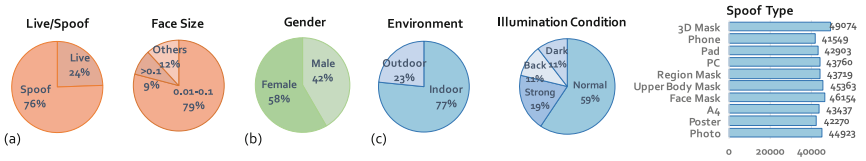

Fig. 3. The statistical distribution of CelebA-Spoof dataset. (a) Overall live and spoof distribution as well as the face size statistic. (b) An exemplar of live attribute, *i.e.* "gender". (c) Three types of spoof attributes

is annotated with 43 attributes. This abundant information enrich the diversity and make face anti-spoofing more illustrative. To our best knowledge, our dataset surpasses all the existing datasets both in scale and diversity.

In this section, we describe our CelebA-Spoof dataset and analyze it through a variety of informative statistics. The dataset is built based on CelebA [20], where all the live people in this dataset are from CelebA. We collect and annotate Spoof images of CelebA-Spoof.

3.1 Semantic Information Collection

In recent decades, studies in attribute-based representations of objects, faces, and scenes have drawn large attention as a complement to categorical representations. However, rare works attempt to exploit semantic information in face anti-spoofing. Indeed, for face anti-spoofing, additional semantic information can characterize the target images by attributes rather than discriminated assignment into a single category, *i.e.* "live" or "spoof".

Semantic for Live - Face Attribute $\mathcal{S}^{\mathbf{f}}$. In our dataset, we directly adopt 40 types of face attributes defined in CelebA [20] as "live" attributes. Attributes of "live" faces always refer to gender, hair color, expression and *etc.* These abundant semantic cues have shown their potential in providing more information for face identification. It is the first time to incorporate them into face anti-spoofing. Extensive studies can be found in Sect. 5.1.

Semantic for Spoof - Spoof Type $\mathcal{S}^{\mathbf{s}}$, Illumination $\mathcal{S}^{\mathbf{i}}$, and Environment $\mathcal{S}^{\mathbf{e}}$. Differs to "live" face attributes, "spoof" images might be characterized by another bunch of properties or attributes as they are not only related to the face region. Indeed, the material of spoof type, illumination condition and environment where spoof images are captured can express more semantic information in "spoof" images, as shown in Fig. 2. Note that the combination of illumination and environment forms the "session" defined in the existing face anti-spoofing dataset. As shown in Table 1, the combination of four illumination conditions and two environments forms 8 sessions. To our best knowledge, CelebA-Spoof is the first dataset covering spoof images in outdoor environment.

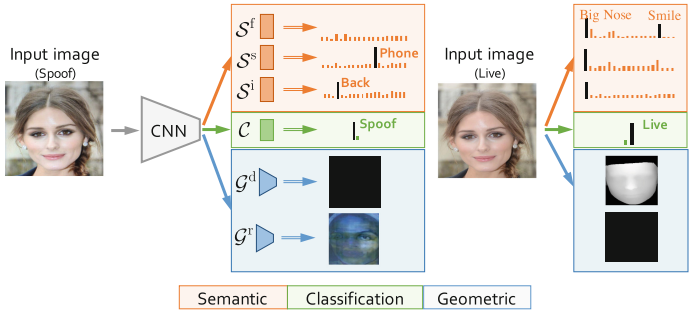

Fig. 4. Auxiliary information Embedding Network (AENet). We use two $\texttt{Conv}_{3\times3}$ after CNN and upsample to size 14×14 to learn the geometric information. Besides, we use three `FC` layers to learn the semantic information. The prediction score of $\mathcal{S}^{\mathrm{f}}$ of spoof image should be very low and the prediction result of $\mathcal{S}^{\mathrm{s}}$ and $\mathcal{S}^{\mathrm{i}}$ of live image should be "No illumination" and "No attack" which belongs to the first label in $\mathcal{S}^{\mathrm{s}}$ and $\mathcal{S}^{\mathrm{i}}$ (Color figure online)

3.2 Statistics on CelebA-Spoof Dataset

The CelebA-Spoof dataset is constructed with a total of 625,537 images. As shown in Fig. 3(a), the ratio of live and spoof is 1 : 3. Face size in all images is mainly between 0.01 million pixels to 0.1 million pixels. We split the CelebA-Spoof dataset into training, validation, and test sets with a ratio of 8 : 1 : 1. Note that all three sets are guaranteed to have no overlap on subjects, which means there is no case of a live image of one certain subject in the training set while its counterpart spoof image in the test set. The distribution of live images in three splits is the same as that defined in the CelebA dataset.

The semantic attribute statistics are shown in Fig. 3(c). The portion of each type of attack is almost the same to guarantee a balanced distribution. It is easy to collect data under normal illumination in an indoor environment where most existing datasets adopt. Besides such easy cases, in CelebA-Spoof dataset, we also involve 12% dark, 11% back, and 19% strong illumination. Furthermore, both indoor and outdoor environments contain all illumination conditions.

4 Auxiliary Information Embedding Network

Equipped with CelebA-Spoof dataset, in this section, we design a simple yet effective network named **A**uxiliary information **E**mbedding **N**etwork (AENet), as shown in Fig. 4. In addition to the main binary classification branch (in green), we **1)** Incorporate the *semantic* branch (in orange) to exploit the auxiliary capacity of rich annotated semantic attributes in the dataset, and **2)** Benchmark the existing *geometric* auxiliary information within this unified multi-task framework.

AENet$_{\mathcal{C},\mathcal{S}}$. Refers to the multi-task jointly learn auxiliary "semantic" attributes and binary "classification" labels. Such auxiliary semantic attributes defined in our dataset provide complement cues rather than discriminated assignment into a single category. The semantic attributes are learned via the backbone network followed by three `FC` layers. In detail, given a batch of n images, based on AENet$_{\mathcal{C},\mathcal{S}}$, we learn live/spoof class $\{\mathcal{C}_k\}_{k=1}^n$ and semantic information, *i.e.* live face attributes $\{\mathcal{S}_k^{\mathrm{f}}\}_{k=1}^n$, spoof type $\{\mathcal{S}_k^{\mathrm{s}}\}_{k=1}^n$ and illumination conditions $\{\mathcal{S}_k^{\mathrm{i}}\}_{k=1}^n$ simultaneously[1]. The loss function of our AENet$_{\mathcal{C},\mathcal{S}}$ is

$$\mathcal{L}_{c,s} = \mathcal{L}_{\mathcal{C}} + \lambda_f \mathcal{L}_{\mathcal{S}^{\mathrm{f}}} + \lambda_s \mathcal{L}_{\mathcal{S}^{\mathrm{s}}} + \lambda_i \mathcal{L}_{\mathcal{S}^{\mathrm{i}}}, \tag{1}$$

where $\mathcal{L}_{\mathcal{S}^{\mathrm{f}}}$ is binary cross entropy loss. $\mathcal{L}_{\mathcal{C}}$, $\mathcal{L}_{\mathcal{S}^{\mathrm{s}}}$ and $\mathcal{L}_{\mathcal{S}^{\mathrm{i}}}$ are softmax cross entropy losses. We set the loss weights $\lambda_f = 1$, $\lambda_s = 0.1$ and $\lambda_i = 0.01$, λ values are empirically selected to balance the contribution of each loss.

AENet$_{\mathcal{C},\mathcal{G}}$. Besides the semantic auxiliary information, some recent works claim some geometric cues such as *reflection map* and *depth map* can facilitate face anti-spoofing. As shown in Fig. 4 (marked in blue), spoof images exhibit even and the flat surfaces which can be easily distinguished by the depth map. The reflection maps, on the other hand, may display reflection artifacts caused by reflected light from flat surface. However, rare works explore their pros and cons.

AENet$_{\mathcal{C},\mathcal{G}}$ also learn auxiliary geometric information in a multi-task fashion with live/spoof classification. Specifically, we concate a `Conv_`3×3 after the backbone network and upsample to 14×14 to output the geometric maps. We denote depth and reflection cues as $\mathcal{G}^{\mathrm{d}}$ and $\mathcal{G}^{\mathrm{r}}$ respectively. The loss function is defined as

$$\mathcal{L}_{c,g} = \mathcal{L}_c + \lambda_d \mathcal{L}_{\mathcal{G}^{\mathrm{d}}} + \lambda_r \mathcal{L}_{\mathcal{G}^{\mathrm{r}}}, \tag{2}$$

where $\mathcal{L}_{\mathcal{G}^{\mathrm{d}}}$ and $\mathcal{L}_{\mathcal{G}^{\mathrm{r}}}$ are mean squared error losses. λ_d and λ_r are set to 0.1. In detail, refer to [11], the ground truth of the depth map of live image is generated by PRNet [9] and the ground truth of the reflection map of the spoof image is generated by the method in [33]. Besides, the ground truth of the depth map of the spoof image and the ground truth of the reflection map of the live images are zero.

5 Ablation Study on CelebA-Spoof

Based on our rich annotations in CelebA-Spoof and the designed AENet, we conduct extensive experiments to analyze semantic information and geometric information. Several valuable observations have been revealed: **1)** We validate that $\mathcal{S}^{\mathrm{f}}$ and $\mathcal{S}^{\mathrm{s}}$ can facilitate live/spoof classification performance greatly. **2)** We analyze the effectiveness of geometric information on different spoof types and find that depth information is particularly sensitive to dark illumination.

[1] Note that we do not learn environments $\mathcal{S}^{\mathrm{e}}$ since we take face image as input where environment cues (*i.e.* indoor or outdoor) cannot provide more valuable information yet illumination influences much.

Table 2. Different settings in ablation study. For Baseline, we use softmax score of $\mathcal{C}$ for classification (a) For AENet$_{\mathcal{S}}$, we use the average softmax score of $\mathcal{S}^{\mathrm{f}}$, $\mathcal{S}^{\mathrm{s}}$ and $\mathcal{S}^{\mathrm{i}}$ for classification. AENet$_{\mathcal{S}^{\mathrm{f}}}$, AENet$_{\mathcal{S}^{\mathrm{s}}}$ and AENet$_{\mathcal{S}^{\mathrm{i}}}$ refer to each single spoof semantic attribute respectively. Based on AENet$_{\mathcal{C},\mathcal{S}}$, w/o $\mathcal{S}^{\mathrm{f}}$, w/o $\mathcal{S}^{\mathrm{s}}$, w/o $\mathcal{S}^{\mathrm{i}}$ mean AENet$_{\mathcal{C},\mathcal{S}}$ discards $\mathcal{S}^{\mathrm{f}}$, $\mathcal{S}^{\mathrm{s}}$ and $\mathcal{S}^{\mathrm{i}}$ respectively. (b) For AENet$_{\mathcal{G}^{\mathrm{d}}}$, we use $\left\|\mathcal{G}^{\mathrm{d}}\right\|_2$ for classification. Based on AENet$_{\mathcal{C},\mathcal{G}}$, w/o $\mathcal{G}^{\mathrm{d}}$, w/o $\mathcal{G}^{\mathrm{r}}$ mean AENet$_{\mathcal{C},\mathcal{G}}$ discards $\mathcal{G}^{\mathrm{d}}$ and $\mathcal{G}^{\mathrm{r}}$ respectively

(a)	Baseline	AENet$_{\mathcal{S}}$	AENet$_{\mathcal{S}^{\mathrm{f}}}$	AENet$_{\mathcal{S}^{\mathrm{s}}}$	AENet$_{\mathcal{S}^{\mathrm{i}}}$	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{f}}$	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{s}}$	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{i}}$	AENet$_{\mathcal{C},\mathcal{S}}$
Live/Spoof	√					√	√	√	√
Face Attribute		√	√				√	√	√
Spoof Type		√		√		√		√	√
Illumination Conditions		√			√	√	√		√

(b)	Baseline	AENet$_{\mathcal{G}^{\mathrm{d}}}$	AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^{\mathrm{r}}$	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{G}^{\mathrm{d}}$	AENet$_{\mathcal{C},\mathcal{G}}$
Live/Spoof	√		√	√	√
Reflection Map				√	√
Depth map		√	√		√

5.1 Study of Semantic Information

In this subsection, we explore the role of different semantic informations annotated in CelebA-Spoof on face anti-spoofing. Based on AENet$_{\mathcal{C},\mathcal{S}}$, we design eight different models in the Table 2(a). The *key* observations are:

Binary Supervision is Indispensable. As shown in Table 3(a), Compared to baseline, AENet$_{\mathcal{S}}$ which only leverages three semantic attributes to do the auxiliary job cannot surpass the performance of baseline. However, as shown in 3(b), AENet$_{\mathcal{C},\mathcal{S}}$ which jointly learns auxiliary semantic attributes and binary classification significantly improves the performance of baseline. Therefore we can infer that even such rich semantic information cannot fully replace live/spoof information. But live/spoof with semantic attributes as auxiliary information can be more effective. This is because the semantic attributes of an image cannot be included completely, and a better classification performance cannot be achieved only by relying on several annotated semantic attributes. However, semantic attributes can help the model pay more attention to cues in the image, thus improving the classification performance of the model.

Semantic Attribute Matters. From Table 3(c), we study the impact of different individual semantic attributes on AENet$_{\mathcal{C},\mathcal{S}}$. As shown in this table, AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{s}}$ achieves the worst APCER. Since APCER reflects the classification ability of spoof images, it shows that compared to other semantic attributes, *spoof types* would significantly affect the performance of the spoof images classification of AENet$_{\mathcal{C},\mathcal{S}}$. Furthermore, we list detail information of AENet$_{\mathcal{C},\mathcal{S}}$ in Fig. 5(a). As shown in this figure, AENet$_{\mathcal{C},\mathcal{S}}$ without *spoof types* gets the 5 worst APCER$_{\mathcal{S}^{\mathrm{s}}}$ out of 10 APCER$_{\mathcal{S}^{\mathrm{s}}}$ and we show up these 5 values in this figure. Besides, in Table 3(b), AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{f}}$ gets the highest BPCER. And we also obtain the BPCER$_{\mathcal{S}^{\mathrm{f}}}$ of each face attribute. As shown in Fig. 5(b), among 40 face attributes, BPCER$_{\mathcal{S}^{\mathrm{f}}}$ of AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{f}}$ occupies 25 worst scores. Since BPCER reflects the classification ability of live images, it demonstrate $\mathcal{S}^{\mathrm{f}}$ plays an important role in the classification of live images.

Qualitative Evaluation. Success and failure cases on live/spoof and semantic attributes predictions are shown in Fig. 6. For *live* examples, the first example

Table 3. Semantic information study results in Sect. 5.1. (a) AENet$_{\mathcal{S}}$ which only depends on semantic attributes for classification cannot surpass the performance of baseline. (b) AENet$_{\mathcal{C},\mathcal{S}}$ which leverages all semantic attributes achieve the best result. **Bolds** are the best results; ↑ means bigger value is better; ↓ means smaller value is better

	Model	Recall (%)↑			AUC ↑	EER (%) ↓	APCER (%) ↓	BPCER (%) ↓	ACER (%) ↓
		FPR = 1%	FPR = 0.5%	FPR = 0.1%					
(a)	Baseline	97.9	95.3	85.9	0.9984	1.6	6.1	1.6	3.8
	AENet$_{\mathcal{S}}$	98.0	96.0	80.4	0.9981	1.4	6.89	1.44	4.17
(b)	AENet$_{\mathcal{C},\mathcal{S}}$	**98.8**	**97.4**	**90.0**	**0.9988**	**1.1**	**4.62**	**1.09**	**2.85**
(c)	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{i}}$	98.1	96.5	86.4	0.9982	1.3	**4.62**	1.35	2.99
	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{s}}$	98.2	96.5	89.4	0.9986	1.3	5.31	1.25	3.28
	AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{f}}$	97.8	95.4	83.6	0.9979	1.3	5.19	1.37	3.28

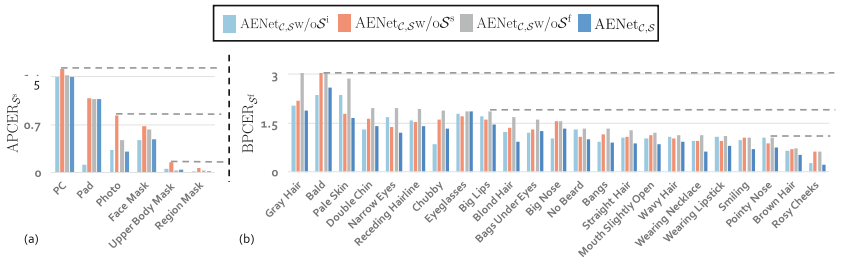

Fig. 5. Representative examples of dropping partial semantic attributes on AENet$_{\mathcal{C},\mathcal{S}}$ performance. In detail, higher APCER$_{\mathcal{S}^{\mathrm{s}}}$ and BPCER$_{\mathcal{S}^{\mathrm{f}}}$ are worse results. (a) Spoof types where AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{s}}$ achieve the worst APCER$_{\mathcal{S}^{\mathrm{s}}}$. (b) Face attributes where AENet$_{\mathcal{C},\mathcal{S}}$ w/o $\mathcal{S}^{\mathrm{f}}$ achieve the worst BPCER$_{\mathcal{S}^{\mathrm{f}}}$

in Fig. 6(a-i) with "glasses" and "hat" help AENet$_{\mathcal{C},\mathcal{S}}$ to pay more attention to the clues of the live image and further improve the performance of prediction of live/spoof. Besides, the first example in Fig. 6(a-ii). AENet$_{\mathcal{C},\mathcal{S}}$ significantly improve the classification performance of live/spoof comparing to baseline. This is because spoof semantic attributes including "back illumination" and "phone" help AENet$_{\mathcal{C},\mathcal{S}}$ recognize the distinct characteristics of spoof image. Note that the prediction of the second example in Fig. 6(b-i) is mistaken.

5.2 Study of Geometric Information

Based on AENet$_{\mathcal{C},\}}$ under different settings, we design four models as shown in Table 2(b) and use semantic attributes we annotated to analyze the usage of geometric information in face anti-spoofing task. The *key* observations are:

Depth Maps are More Versatile. As shown in Table 4(a), geometric information is insufficient to be the unique supervision for live/spoof classification. However, it can boost the performance of the baseline when it serves as an auxiliary supervision. Besides, we study the impact of different individual geometric information on AENet$_{\mathcal{C},\mathcal{G}}$ performance. As shown in Fig. 7(a), AENet$_{\mathcal{C},\mathcal{G}}$

Table 4. Geometric information study results in Sect. 5.2. (a) AENet$_{\mathcal{G}^d}$ which only depends on the depth map for classification performs worst than baseline. (b) AENet$_{\mathcal{C},\mathcal{G}}$ which leverages all semantic attributes achieve the best result. **Bolds** are the best results; ↑ means bigger value is better; ↓ means smaller value is better

	Model	Recall (%)↑			AUC↑	EER (%)↓	APCER (%)↓	BPCER (%)↓	ACER (%)↓
		FPR = 1%	FPR = 0.5%	FPR = 0.1%					
(a)	Baseline	97.9	95.3	85.9	0.9984	1.6	6.1	1.6	3.8
	AENet$_{\mathcal{G}^d}$	97.8	96.2	87.0	0.9946	1.6	7.33	1.68	4.51
(b)	AENet$_{\mathcal{C},\mathcal{G}}$	**98.4**	**96.8**	86.7	**0.9985**	**1.2**	**5.34**	**1.19**	**3.26**
(c)	AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^d$	98.3	96.1	**87.7**	0.9976	**1.2**	5.91	1.27	3.59
	AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$	97.9	95.7	84.1	0.9973	1.3	5.71	1.38	3.55

w/o $\mathcal{G}^d$ performs the best in spoof type: "replay" (macro definition), because the reflect artifacts appear frequently in these three spoof types. For "phone", AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^d$ improves 56% comparing to the baseline. However AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^d$ gets worse result than baseline in spoof type: "print" (macro definition). Moreover, AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$ helps greatly to improve the classification performance of baseline in both "replay" and "print"(macro definition). Especially for "poster", AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$ improves baseline by 81%. Therefore, the depth map can improve classification performance in most spoof types, but the function of the reflection map is mainly reflected in "replay"(macro definition).

Sensitive to Illumination. As shown in Fig. 7(a), in spoof type "print"(macro definition), the performance of the AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$ on "A4" is much worse than "poster" and "photo", although they are both in "print" spoof type. The main reason for the large difference in performance among these three spoof types for AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$ is that the learning of the depth map is sensitive to dark illumination, as shown in Fig. 7(b). When we calculate APCER under other illumination conditions: normal, strong and back, AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^r$ achieves almost the same results among "A4", "poster" and "photo".

6 Benchmarks

In order to facilitate future research in the community, we carefully build three different benchmarks to investigate face anti-spoofing algorithms. Specifically, for a comprehensive evaluation, besides ResNet-18, we also provide the corresponding results based on a heavier backbone, *i.e.* Xception. Detailed information of the results based on Xception are shown in the supplementary material.

6.1 Intra-dataset Benchmark

Based on this benchmark, models are trained and evaluated on the whole training set and testing set of CelebA-Spoof. This benchmark evaluates the overall capability of the classification models. According to different input data types, there are two kinds of face anti-spoof methods, *i.e.* " video-driven methods"

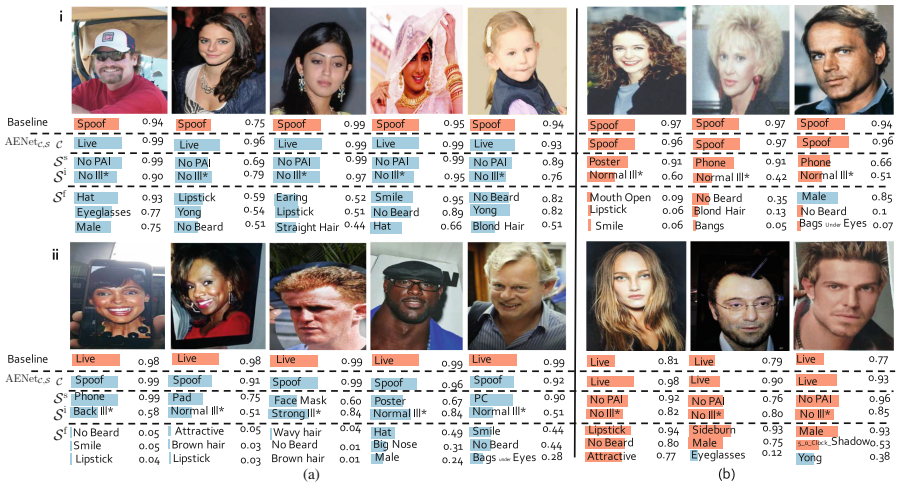

Fig. 6. Success and failure cases. The row(i) present the live image and row(ii) present the spoof image. For each image, the first row is the highest score of live/spoof prediction of baseline and others are the highest live/spoof and the highest semantic attributes predictions of AENet$_{\mathcal{C},\mathcal{S}}$. Blue indicates correctly predicted results and orange indicates the wrong results. In detail, we list the top three prediction scores of face attributes in the last three rows of each image (Color figure online)

Table 5. Intro-dataset Benchmark results on CelebA-Spoof. AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$ achieved the best result. **Bolds** are the best results; ↑ means bigger value is better; ↓ means smaller value is better. * Model 2 defined in *Auxiliary* can be used as "image driven method"

Model	Backbone	Parm. (MB)	Recall (%)↑			AUC↑	EER (%)↓	APCER (%)↓	BPCER (%)↓	ACER (%)↓
			FPR = 1%	FPR = 0.5%	FPR = 0.1%					
Auxiliary* [18]	–	22.1	97.3	95.2	83.2	0.9972	1.2	5.71	1.41	3.56
BASN [11]	VGG16	569.7	98.9	**97.8**	**90.9**	**0.9991**	1.1	4.0	1.1	2.6
AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$	ResNet-18	42.7	**98.9**	97.3	87.3	0.9989	**0.9**	**2.29**	**0.96**	**1.63**

and "image-driven methods". Since the data in CelebA-Spoof are image-based, we benchmark state-of-the-art "image-driven methods" in this subsection. As shown in Table 5, AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$ which combines geometric and semantic information has achieved the best results on CelebA-Spoof. Specifically, our approach outperforms the state-of-the-art by 38% with much fewer parameters.

6.2 Cross-Domain Benchmark

Since face anti-spoofing is an open-set problem, even though CelebA-Spoof is equipped with diverse images, it is impossible to cover all spoof types, environments, sensors, *etc.* that exist in the real world. Inspired by [4,18], we carefully design two protocols for CelebA-Spoof based on real-world scenarios. In each protocol, we evaluate the performance of trained models under controlled domain

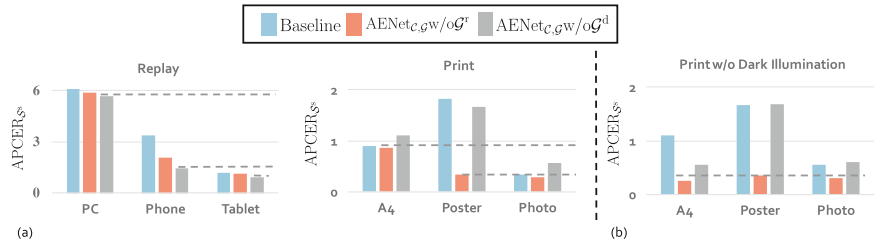

Fig. 7. Representative examples of the effectiveness of geometric information. Higher APCER$_{\mathcal{S}^s}$ is worse. (a) AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^{d}$ perform the best in spoof type: "replay"(macro definition) and AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^{r}$ perform the best in spoof type: "print"(macro definition). (b) The performance of AENet$_{\mathcal{C},\mathcal{G}}$ w/o $\mathcal{G}^{r}$ improve largely on spoof type: "A4", if we only calculate APCER under illumination conditions: "normal", "strong" and "back"

Table 6. Cross-domain benchmark results of CelebA-Spoof. **Bolds** are the best results; ↑ means bigger value is better; ↓ means smaller value is better

Protocol	Model	Recall (%) ↑			AUC↑	EER (%)↓	APCER (%)↓	BPCER (%)↓	ACER (%)↓
		FPR = 1%	FPR = 0.5%	FPR = 0.1%					
1	Baseline	93.7	86.9	69.6	**0.996**	2.5	5.7	2.52	4.11
	AENet$_{\mathcal{C},\mathcal{G}}$	93.3	88.6	**74.0**	0.994	2.5	5.28	2.41	3.85
	AENet$_{\mathcal{C},\mathcal{S}}$	93.4	89.3	71.3	**0.996**	2.4	5.63	2.42	4.04
	AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$	**95.0**	**91.4**	73.6	0.995	**2.1**	**4.09**	**2.09**	**3.09**
2	Baseline	#	#	#	0.998 ±0.002	1.5±0.8	8.53±2.6	1.56±0.81	5.05±1.42
	AENet$_{\mathcal{C},\mathcal{G}}$	#	#	#	0.995±0.003	1.6±4.5	8.95±1.07	1.67±0.9	5.31±0.95
	AENet$_{\mathcal{C},\mathcal{S}}$	#	#	#	0.997±0.002	1.2±0.7	**4.01±2.9**	1.24±0.67	3.96±1.79
	AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$	#	#	#	**0.998±0.002**	**1.3±0.7**	4.94±3.42	**1.24±0.73**	**3.09±2.08**

shifts. Specifically, we define two protocols. **1)** *Protocol 1* - Protocol 1 evaluates the cross-medium performance of various spoof types. This protocol includes 3 macro types of spoof, where each covers 3 micro types of spoof. These three macro types of spoof are "print", "repay" and "paper cut". In detail, in each macro type of spoof, we choose 2 of their micro type of spoof for training, and the others for testing. Specifically, "A4", "face mask" and "PC" are selected for testing. **2)** *Protocol 2* - Protocol 2 evaluates the effect of input sensor variations. According to imaging quality, we split input sensors into three groups: *low-quality sensor*, *middle-quality sensor* and *high-quality sensor*[2]. Since we need to test on three different kinds of sensor and the average performance of FPR-Recall is hard to measure, we do not include FPR-Recall in the evaluation metrics of protocol 2. Table 6 shows the performance under each protocol.

6.3 Cross-Dataset Benchmark

In this subsection, we perform cross-dataset testing on CelebA-Spoof and CASIA-MFSD dataset to further construct the cross-dataset benchmark. On the one hand, we offer a quantitative result to measure the quality of our

[2] Please refer to supplementary for the detailed input sensors information.

Table 7. Cross-dataset benchmark results. AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$ based on ResNet-18 achieves the best generalization performance. **Bolds** are the best results; ↑ means bigger value is better; ↓ means smaller value is better

Model	Training	Testing	HTER (%) ↓
FAS-TD-SF [28]	SiW	CASIA-MFSD	39.4
FAS-TD-SF [28]	CASIA-SURF	CASIA-MFSD	37.3
AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$	SiW	CASIA-MFSD	27.6
Baseline	CelebA-Spoof	CASIA-MFSD	14.3
AENet$_{\mathcal{C},\mathcal{G}}$	CelebA-Spoof	CASIA-MFSD	14.1
AENet$_{\mathcal{C},\mathcal{S}}$	CelebA-Spoof	CASIA-MFSD	12.1
AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$	CelebA-Spoof	CASIA-MFSD	**11.9**

dataset. On the other hand, we can evaluate the generalization ability of different methods according to this benchmark. The current largest face anti-spoofing dataset CASIA-SURF [32] adopted *FAS-TD-SF* [28] (which is trained on SiW or CASIA-SURF and tested on CASIA-MFSD) to demonstrate the quality of CASIA-SURF. Following this setting, we first train AENet$_{\mathcal{C},\mathcal{G}}$, AENet$_{\mathcal{C},\mathcal{S}}$ and AENet$_{\mathcal{C},\mathcal{S},\mathcal{G}}$ based on CelebA-Spoof and then test them on CASIA-MFSD to evaluate the quality of CelebA-Spoof. As shown in Table 7, we can conclude that: **1)** The diversity and large quantities of CelebA-Spoof drastically boosts the performance of vanilla model; a simple ResNet-18 achieves state-of-the-art cross-dataset performance. **2)** Comparing to geometric information, semantic information equips the model with better generalization ability.

7 Conclusion

In this paper, we construct a large-scale face anti-spoofing dataset, **CelebA-Spoof**, with 625,537 images from 10,177 subjects, which includes 43 rich attributes on face, illumination, environment and spoof types. We believe CelebA-Spoof would be a significant contribution to the community of face anti-spoofing. Based on these rich attributes, we further propose a simple yet powerful multi-task framework, namely **AENet**. Through AENet, we conduct extensive experiments to explore the roles of semantic information and geometric information in face anti-spoofing. To support comprehensive evaluation and diagnosis, we establish three versatile benchmarks to evaluate the performance and generalization ability of various methods under different carefully-designed protocols. With several valuable observations revealed, we demonstrate the effectiveness of CelebA-Spoof and its rich attributes which can significantly facilitate future research.

Acknowledgments. This work is supported in part by SenseTime Group Limited, in part by National Science Foundation of China Grant No. U1934220 and 61790575, and the project "Safety data acquisition equipment for industrial enterprises No.134". The corresponding author is Jing Shao. The contributions of Yuanhan Zhang and Zhenfei Yin are Equal.

References

1. Bhattacharjee, S., Mohammadi, A., Marcel, S.: Spoofing deep face recognition with custom silicone masks. In: Proceedings of IEEE 9th International Conference on Biometrics: Theory, Applications, and Systems (BTAS) (2018)
2. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face antispoofing using speeded-up robust features and fisher vector encoding. IEEE Signal Process. Lett. **24**(2), 141–145 (2016)
3. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face spoofing detection using colour texture analysis. TIFS **11**(8), 1818–1830 (2016)
4. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: Oulu-npu: a mobile face presentation attack database with real-world variations. In: FG, pp. 612–618. IEEE (2017)
5. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: BIOSIG, pp. 1–7. IEEE (2012)
6. Chingovska, I., Erdogmus, N., Anjos, A., Marcel, S.: Face recognition systems under spoofing attacks. In: Bourlai, T. (ed.) Face Recognition Across the Imaging Spectrum, pp. 165–194. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28501-6_8
7. Erdogmus, N., Marcel, S.: Spoofing 2D face recognition systems with 3D masks. In: BIOSIG, pp. 1–8. IEEE (2013)
8. Feng, L., Po, L.M., Li, Y., Xu, X., Yuan, F., Cheung, T.C.H., Cheung, K.W.: Integration of image quality and motion cues for face anti-spoofing: a neural network approach. J. Visual Commun. Image Represent. **38**, 451–460 (2016)
9. Feng, Y., Wu, F., Shao, X., Wang, Y., Zhou, X.: Joint 3D face reconstruction and dense alignment with position map regression network. In: ECCV, pp. 534–551 (2018)
10. Jourabloo, A., Liu, Y., Liu, X.: Face de-spoofing: anti-spoofing via noise modeling. In: ECCV, pp. 290–306 (2018)
11. Kim, T., Kim, Y., Kim, I., Kim, D.: BASN: enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In: ICCV Workshops (2019)
12. Kollreider, K., Fronthaler, H., Faraj, M.I., Bigun, J.: Real-time face detection and motion analysis with application in "liveness" assessment. TIFS **2**(3), 548–558 (2007)
13. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of fourier spectra. In: Biometric Technology for Human Identification, vol. 5404, pp. 296–303. International Society for Optics and Photonics (2004)
14. Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A.: An original face anti-spoofing approach using partial convolutional neural network. In: IPTA, pp. 1–6. IEEE (2016)
15. Liu, S.Q., Lan, X., Yuen, P.C.: Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection. In: ECCV, September 2018

16. Liu, S., Yang, B., Yuen, P.C., Zhao, G.: A 3D mask face anti-spoofing database with real world variations. In: CVPR Workshops, pp. 1551–1557, June 2016
17. Liu, S., Yuen, P.C., Zhang, S., Zhao, G.: 3D mask face anti-spoofing with remote photoplethysmography. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 85–100. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_6
18. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: binary or auxiliary supervision. In: CVPR, pp. 389–398 (2018)
19. Liu, Y., Stehouwer, J., Jourabloo, A., Liu, X.: Deep tree learning for zero-shot face anti-spoofing. In: CVPR, pp. 4680–4689 (2019)
20. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: ICCV (2015)
21. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using texture and local shape analysis. IET Biom. **1**(1), 3–10 (2012)
22. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. TPAMI **24**(7), 971–987 (2002)
23. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: ICCV, pp. 1–8. IEEE (2007)
24. Patel, K., Han, H., Jain, A.K.: Secure face unlock: spoof detection on smartphones. TIFS **11**(10), 2268–2283 (2016)
25. Schwartz, W.R., Rocha, A., Pedrini, H.: Face spoofing detection through partial least squares and low-level descriptors. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8. IEEE (2011)
26. Shao, R., Lan, X., Li, J., Yuen, P.C.: Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: CVPR (2019)
27. Sun, L., Pan, G., Wu, Z., Lao, S.: Blinking-based live face detection using conditional random fields. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 252–260. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74549-5_27
28. Wang, Z., et al.: Exploiting temporal and depth information for multi-frame face anti-spoofing. arXiv (2018)
29. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. TIFS **10**(4), 746–761 (2015)
30. Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection with component dependent descriptor. In: ICB, pp. 1–6. IEEE (2013)
31. Yang, X., et al.: Face anti-spoofing: model matters, so does data. In: CVPR, pp. 3507–3516 (2019)
32. Zhang, S., et al.: A dataset and benchmark for large-scale multi-modal face anti-spoofing. In: CVPR, pp. 919–928 (2018)
33. Zhang, X., Ng, R., Chen, Q.: Single image reflection separation with perceptual losses. In: ICCV, pp. 4786–4794 (2018)
34. Zhang, Z., et al.: A face antispoofing database with diverse attacks. In: ICB, pp. 26–31. IEEE (2012)